REVIEW

Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations

BUYONG MA,[1] MAXIM SHATSKY,[2] HAIM J. WOLFSON,[2] AND RUTH NUSSINOV[3,4]

[1]Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick, Frederick, Maryland 21702, USA
[2]School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[3]Intramural Research Support Program–SAIC, National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Experimental and Computational Biology, Frederick, Maryland 21702, USA
[4]Sackler Institute of Molecular Medicine, Department of Human Genetics, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

(RECEIVED May 31, 2001; FINAL REVISION October 12, 2001; ACCEPTED November 1, 2001)

Abstract

Here, we comment on the steadily increasing body of data showing that proteins with specificity actually bind ligands of diverse shapes, sizes, and composition. Such a phenomenon is not surprising when one considers that binding is a dynamic process with populations in equilibrium and that the shape of the binding site is strongly influenced by the molecular partner. It derives implicitly from the concept of populations. All proteins, specific and nonspecific, exist in ensembles of substates. If the library of ligands in solution is large enough, favorably matching ligands with altered shapes and sizes can be expected to bind, with a redistribution of the protein populations. Point mutations at spatially distant sites may exert large conformational rearrangements and hinge effects, consistent with mutations away from the binding site leading to population shifts and (cross-)drug resistance. A similar effect is observed in protein superfamilies, in which different sequences with similar topologies display similar large-scale dynamic motions. The hinges are frequently at analogous sites, yet with different substrate specificity. Similar topologies yield similar conformational isomers, although with different distributions of population times, owing to the change in the conditions, that is, the change in the sequences. In turn, different distributions relate to binding of different sizes and shapes. Hence, the binding site shape and size are defined by the ligand. They are not independent entities of fixed proportions and cannot be analyzed independently of the binding partner. Such a proposition derives from viewing proteins as dynamic distributions, presenting to the incoming ligands a range of binding site shapes. It illustrates how presumably specific binding molecules can bind multiple ligands. In terms of drug design, the ability of a single receptor to recognize many dissimilar ligands shows the need to consider more diverse molecules. It provides a rationale for higher affinity inhibitors that are not derived from substrates at their transition states and indicates flexible docking schemes.

Keywords: Multiple ligand binding; single site binding; hinge bending; combinatorial libraries; flexible docking; populations; flexible structural comparison

Until fairly recently, proteins and their binding sites have been divided into two general categories: specific versus broad-range binding proteins, such as proteolytic enzymes or germ line antibodies (for review, see Ma et al. 1999). With regard to specific binding, a major question has long been what determines ligand recognition (Ringe 1995). For the less selective, broader-range binding, the question arises as to why different ligands still bind at the same single site rather than at different locations on the protein surface (DeLano et al. 2000; Ma et al. 2001).

Reprints request to Ruth Nussinov, NCI-FCRDC Bldg 469, Room 151, Frederick, MD 21702; email: ruthn@ncifcrf.gov; fax: (301) 846-5598.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.21302.

Recently, through utilization of combinatorial libraries, a particularly intriguing observation has been made. First, for the broader-range ligand binding proteins, it has been observed that ligands with composition, sizes, and shapes that are different from those of the natural-binding protein ligands may bind with an equal or even higher affinity at the same site (DeLano et al. 2000; Vazquez-Laslop et al. 2000; Zwahlen et al. 2000). Further, even presumably specific enzymes or receptors may bind ligands of different shape, size, and composition, with (sometimes) higher affinities (Tondi et al. 1999). Here we derive some of the implications with regard to both the determinants of binding sites and inhibitor design. We argue that both types of cases, specific and multiple-ligand binding at a single site, are the outcome of the distribution of conformational isomers, with variable extents of largely hinge-bending movements that exist around the native state. Currently, synthetic inhibitors largely mimic natural substrates and are frequently transition state analogs. However, if the binding range is substantially broader, better-fitting inhibitors with higher affinities can potentially be designed (Tondi et al. 1999).

The principles of binding are universal and apply equally well to both selective and multiple-ligand single-site binding. Hence, it is not unreasonable to expect that as additional specific-binding proteins are analyzed through in vitro combinatorial library selection, a substantial fraction will bind to ligands of altered shape, size, and composition. Multiple-ligand binding at a single site simply reflects the existence of populations of conformers in solution. The distribution of conformers in the ensemble is largely the outcome of the protein topology (Sinha and Nussinov 2001), modulated by the types of residues at certain locations and by conditions such as the presence of ligands (Tsai et al. 1999; Carlson and McCammon 2000; Kumar et al. 2000). Mutations and binding to given ligands/substrates are likely to lead to population shifts in pre-existing conformers.

Binding is a process, not a still-life jigsaw interlocking of two molecules. Side-chain movements are invariably involved in all protein binding processes. However, hinge-bending motions are also likely to universally exist, even though they might not be observed experimentally in a bound versus unbound structural comparison. Different conformers in solution may shift to a practically identical conformation observed in the crystal. Hinge bending reflects a range of conformations separated by low-energy barriers (Gerstein et al. 1994; Gerstein and Krebs 1998). Binding multiple ligands with different sizes and shapes implies hinge-type motions. However, proteins considered to display specificity also exist in an ensemble of conformers, with their domains (subdomains) rotated and bent with respect to each other. Should the solution contain a ligand fitting a hinge-bent conformation with a favorable geometry and chemistry, that is, with its free energy greater than the free energy difference between the two protein states (Carlson and McCammon 2000), the ligand would bind at that same site. Even if the favorable protein conformer has a low population time, binding to its ligand will shift the equilibrium in its favor, further driving the binding reaction (Kumar et al. 2000) and changing the average structure of the protein (Carlson and McCammon 2000). Hinge-bending motions may expand the binding site or contract it, yielding different sizes and shapes. Figure 1 presents a schematic diagram of this concept. The protein illustrated is the tissue factor (Banner et al. 1996). Tissue factor is in complex with blood coagulation factor VIIa with the binding sites located at the hinge region (three residues on one domain, i.e., R131, L133, and F140, and six residues on the other domain, i.e., T17, K20, I22, E24, E56, and D58; Banner et al. 1996). In room temperature, molecular dynamics simulations using CHARMm EEF1 forcefield (Brooks et al. 1983; Lazaridis and Karplus 1999), we have observed that the relative orientation of the two binding patches changes with the hinge motion significantly. Figure 1, A and B, is a simplification of the motion observed in the simulations. Different ligands selectively bind to different tissue conformers.

Specific versus multiple-ligand binding at a single site: A matter of pre-existing populations

Van Regenmortel (1999) has recently remarked that the structure of a binding site cannot be visualized without consideration of the binding partner. Around the native state, the protein exists in a range of conformations. In the process of binding, the conformer that is selected is the one with a binding site that complements most favorably that of the incoming ligand (Ma et al. 1999). Binding-site selection is ligand dependent (Ma et al. 2001). Hence, as Van Regenmortel has argued, binding site shape and size are defined by the ligand. They are not independent entities of fixed proportions and cannot be analyzed independently of their partner.

Such a view is consistent with dynamic populations, with low-energy barriers between them. In all conformers, the structure of the molecule is around the native state. However, owing to the location of the binding site, frequently consisting of parts that move with respect to each other, the molecular ensemble illustrates binding sites with different geometries. The snapshot geometry captured in the unbound crystal structure is just one out of many, corresponding to visual time slices of dynamic systems (Van Regenmortel 1999). The crystal structure represents a weighted average of the ensemble, and (depending on the landscape) it could correspond to a state with a higher free energy.

Viewing molecules as dynamic distributions, presenting to the incoming ligands a range of binding site shapes, illustrates how presumably specific binding molecules can bind multiple ligands. This happens in cases in which the

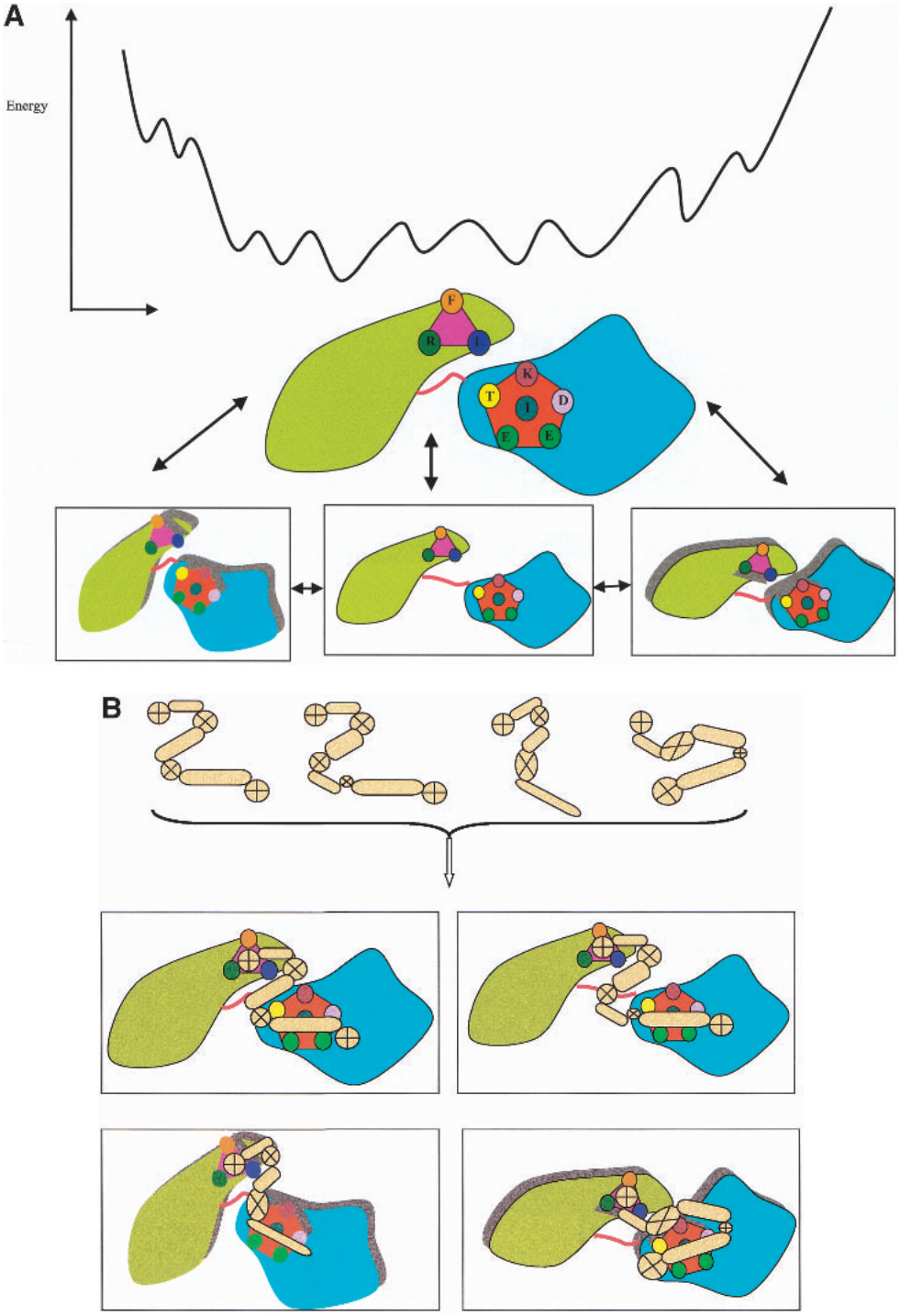

Fig. 1. The concept described in this mini-review. Around the native state, there is a range of conformational isomers, with low barriers separating between them. These conformational isomers largely reflect large-scale motions, an outcome of hinge-bending type movements. The geometry of the binding site reflects this variability. Depending on the library of ligands present in solution and on the conditions, the protein may selectively bind to these, still at the same binding site. The protein illustrated here is the tissue factor (Banner et al. 1996). The tissue factor is in complex with blood coagulation factor VIIa with the binding sites located in the hinge region (three residues on one domain, i.e., R131, L133, and F140, and six residues on another domain, i.e., T17, K20, I22, E24, E56, and D58). This figure is a simplification of the motion observed in the molecular dynamics simulations (also see Fig. 4). (*A*) The energy landscape of this protein, with the minima corresponding to different hinge-bending tissue factor conformers. The conformers dynamically interconvert in solution. Arrows are drawn connecting minima in the funnel and different respective hinge-bent conformers. (*B*) Each such conformer is bound to a different ligand. The different ligands present in solution are drawn at the top of the figure. On mixing with the tissue conformers, the ligands selectively choose different tissue factor conformers, shifting the equilibrium in favor of these conformers.

ligand library is large enough and includes well-fitting molecules. It further illustrates why multiple ligands still preferentially bind at a single site (DeLano et al. 2000).

An excellent example relates to the HIV protease and several aspartic proteases, including pepsin (Abad-Zapatero et al. 1990), human cathepsin D (Baldwin et al. 1993), and plasmepsin II (Silva et al. 1996), an enzyme isolated from *Plasmodium falciparum*. Plasmepsin II is essential for the proteolytic digestion of hemoglobin, which takes place in digestive vacuoles of malarial parasites, and is a target of antimalarial drug design. Silva et al. (1996) have determined the crystal structure of recombinant plasmepsin II. The space group contained two independent proteins in the asymmetric unit. Remarkably, the two proteins have displayed different domain displacements (Lee et al. 1998). Hence, here what Lee et al. have shown is a direct and striking example of domain flexibility within a single crystal form, indicating that the enzyme must pre-exist in a dynamic equilibrium between at least two observed states. This is even more remarkable, considering that the proteins were complexed with pepstatin A, a general aspartic proteinase inhibitor and a potent inhibitor of plasmepsin II. More recent results by Silva et al. on a second inhibitor complex of plasmepsin II have illustrated still a third domain displacement (Erickson, personal comm.).

Tissue factor provides an additional example. There, two proteins also in the same crystallographic asymmetry unit differ in their relative domain-domain orientation. Muller et al. (1998) have crystallized the tissue factor, a member of the cytokine receptor superfamily. Tissue factor is an obligate cofactor of coagulation factor VIIa. Muller et al. compared the two molecules in the asymmetric unit and showed that there is a hinge rotation of 12.7° in the rabbit tissue factor between the domains. This indicates that under crystallization conditions, at least one of these conformers was in a highly populated state, with another conformer complementary to it favorably binding and co-crystallizing with it. Interestingly, Huang et al. (1998) have observed a 7° hinge-bending angle at the same location in the unbound human tissue factor (h-TF) compared with h-TF in complex with an antigen-binding fragment (Fab). In the plasmepsin case, the different conformers of the complex bind as a unit to the growing crystal. In the tissue factor case, it is different conformers of the single molecules that bind to each other in the asymmetric unit of the crystal. Figure 2 illustrates the hinge bending of tissue factor, using an efficient, entirely automated, hinge-bending flexible structural comparison algorithm (Shatsky et al. 2000; Shatsky 2001).

Proteolytic enzymes catalyze the hydrolysis of the protein backbone via nucleophilic attack on the carbonyl carbon of the peptide bond. During this nucleophilic attack, the trigonal carbonyl carbon of the peptide bond becomes tetragonal in the intermediate, or transition, state (Creighton 1993). There are different genera of proteolytic enzymes (e.g., with serine, threonine, cysteine, aspartic, or metallo groups providing a nucleophile during catalysis), with extensive structural diversity. For a particular proteinase, only a limited extent of substrate specificity is found. Trypsin prefers to cut the peptide bond following Arg or Lys residues, whereas chymotrypsin cuts the peptide bonds formed by Tyr, Phe, Trp, and Leu. Yet, these enzymes recognize the different substrate conformations that inherently exist in unrelated proteins. Similarly, topoisomerases I and II cleave DNA strands nonspecifically, complementing many conformations resulting from different nucleotide sequences. Here the enzyme recognizes different substrate conformations.

Nevertheless, geometry, although critical for favorable binding, is insufficient. The presence of well-fitting shapes is complemented by polar residue hot spots, present on the protein surface in naturally occurring binding sites (Clackson and Wells 1995, Bogan and Thorn 1998; Clackson et al. 1998; Hu et al. 2000). As the binding site undergoes dynamic changes, the network of hydrogen bonds connecting these residues changes, with different side-chains or backbone polar groups mediating their interactions both within the binding site and with the binding partner across the intermolecular interface (Kranz and Hall 1999).

Multiple-ligand binding at a single site and hinge motions

Specificity is relative. It is determined both by the extent of the hinge-bending flexibility and, hence, by the range of hinge-based binding site shapes and volumes and by the accessible ligands. The role of structural plasticity in protein-protein interactions has been recently reviewed (Sundberg and Mariuzza 2000; Demchenko 2001).

Several pieces of evidence combine to implicate hinge bending in altered binding site shape and size. First, many mutations, perhaps the majority of those conferring drug resistance, occur far from the binding site (Rose et al. 1998). Second, mutations occurring at interfacial interdomain boundaries have been implicated in domain (subdomain) movements. These domains (or parts) move as relatively rigid bodies with respect to each other. Third, domain (subdomain) movements have long been known to be involved in changes in the active site (for examples, see Janin and Wodak 1983; Huber 1987; Joseph et al. 1990; Schulz et al. 1990; Sali et al. 1992; Sharff et al. 1992; Gerstein et al. 1993a,b; Oh et al. 1993; Cox et al. 1994; Mattevi et al. 1996; Wedemyer et al. 1997; Horovitz 1998). Indeed, it suffices to consider the open versus the closed conformation proposed by Koshland (1958) and shown in a large number of cases. Fourth, it has been noted for many cases, ranging from HIV-1 infection to diseases such as cancer, that often naturally occurring mutations selected to combat one drug will confer resistance to some others. Fifth, drugs are generally engineered to have a tightly fitting interface with the

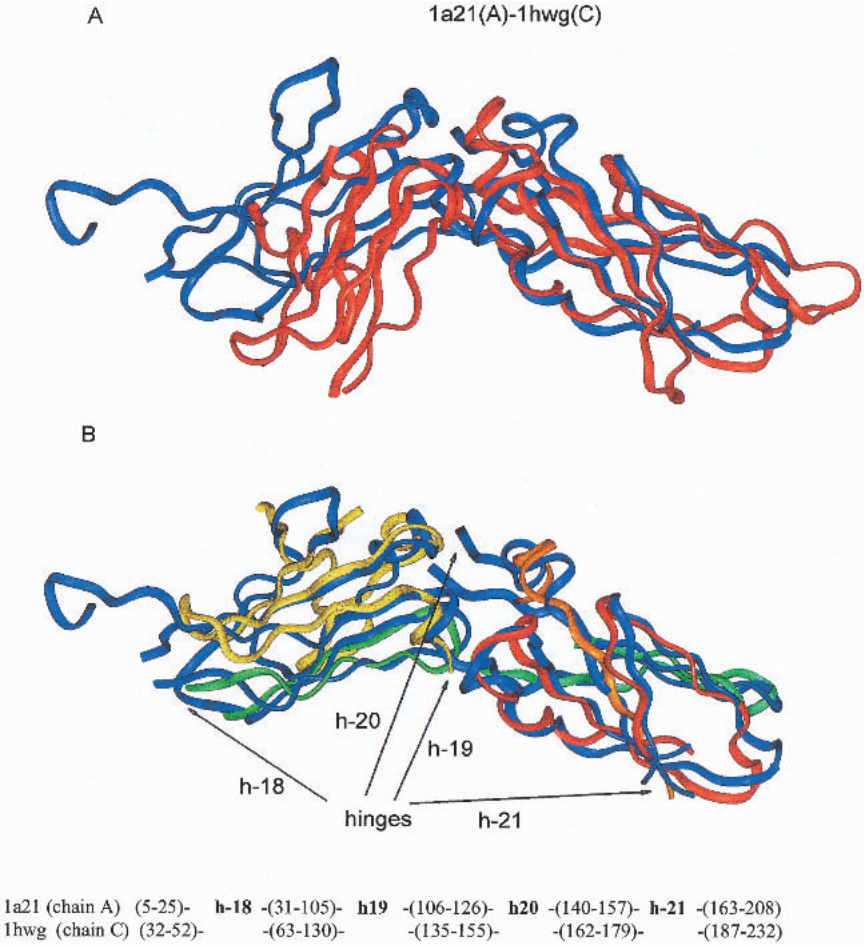

Fig. 2. The hinge bending of tissue factor, using an efficient, entirely automated, hinge-bending flexible structural comparison algorithm (Shatsky et al. 2000; Shatsky 2001). (*A*) A rigid superposition of the tissue factor 1a21 (chain A) and growth hormone binding protein 1hwg (chain C). According to SCOP (Murzin et al. 1995), both proteins belong to the fibronectin type III protein family. The proteins have two domains, each composed of two β-sheets. The sequence similarity between the proteins is minor. The coordinates are taken from the Protein Data Bank (Bernstein et al. 1977). (*B*) Flexible structural superposition. Four hinges are observed. These are labeled on the figure. Each rigid fragment is depicted in a different color. The superposition has been performed by FlexProt. The rigid fragment pairs and the flexible hinge region are detected simultaneously. No predefinition of the hinge sites is required. The program can be accessed at http://bioinfo3d.math.tau.ac.il/FlexProt.

protein active site, with specific favorable interactions. Hence, this cross-drug resistance may well imply that it is not necessarily the mutations at the receptor-ligand binding interface that are solely responsible for the cross resistance (Rose et al. 1998). Straightforward reasonable alternatives are changes in the binding site size, shape, and epitope, hampering the highly favorable drug-receptor binding. Depending on either the interactions at the interfaces between the domains or on the residues at, or near, the hinge site (Sun and Sampson 1998), the domain may move to a greater or lesser extent, altering the range of orientations with respect to each other. Mutations at the interdomain interface, or in the vicinity of the hinge site, may therefore lead to a change in the interdomain orientation. Together, these pieces weave a logical scheme. Drug-resistant mutations occurring far from the binding site may exert their effect via rigid-body domain motions.

These points have been very attractively shown in retroviral proteases, specifically with regard to structural implications for drug design. Rose et al. (1998) have found that rigid body rotation of five domains and movements within their interfacial joints provide a rational context for understanding why HIV protease mutations that arise in drug-resistant strains are often spatially removed from the drug or substrate binding sites. They have identified and characterized domain motions associated with substrate binding in the retroviral HIV-1 and apo-simian immunodeficiency virus (SIV) proteases. These motions are in addition to closure of the flaps and result from rotations of 6° to 7° at primarily hydrophobic interfaces. The crystal structure of the unli-

ganded SIV protease is in the most open conformation of any retroviral protease determined to date. Comparisons of this structure, and of unliganded HIV structures with their corresponding liganded complexes, have illustrated that five domains of the protease dimer move as rigid bodies with respect to one another. These five domains include a terminal domain of the dimer (containing the N- and C-terminal β-sheet of the dimer); two core domains, which contain the catalytic aspartic residues; and two flap domains. Rose et al. (1998) have shown that the two core domains rotate toward each other, reshaping the binding pocket. Further, they have shown that mutations at the interdomain interfaces that favor the unliganded form increase the off-rate of the inhibitor, allowing the substrate greater access for catalysis. This indicates a potential mechanism of resistance to competitive inhibitors, especially when the forward enzymatic reaction rate exceeds the rate of substrate dissociation.

Hence, the cause is the mutations, the end effect is the drug resistance, and the apparent mechanism is the change in the geometry of the binding site, via hinge-bending domain motions.

Molecular ensembles

The conformational space is not equally populated. Some regions are more populated than others, and these are largely determined by the protein topology. Topology determines both the hinge-bending motions (Keskin et al. 2000) and the smaller scale motions, such as observed among point-mutants (Sinha and Nussinov 2001). Comparison of molecules with different point mutations illustrates that the regions of larger deviations are, in general, not at the locations of the mutations. Rather, regardless of the positions of the mutations in the structures, the regions that move the most are the same. The effect of mutations is in redistributions of the populations. Analysis of topologically related structures has also indicated that the hinge-bending motions are at similar locations. Depending on the distributions of the conformers and their conformational variability, they provide a gradient of binding site conformations of different shapes and volumes. These can potentially favorably interact with ensembles of ligands of variable sizes. This gradient has the potential to favorably interact with the ensembles of many diverse ligands, driving the complementary interactions between the two. This induced complementarity would create very different specific interactions for a wide variety of ligands. Such a gradient, along with the available ligands may illustrate specific through multiple-ligand single-site binding. Although ligands largely bind pre-existing conformations, some movements of backbone, and particularly of side-chains, may be induced to optimize the intermolecular interactions (Kumar et al. 2000). Hence, the observation that presumably specific proteins may bind different ligands, is not surprising, given a large enough library of ligands. It falls out of considerations of ensembles and distributions.

Recently, Kern and her colleagues (Volkman et al. 2001) have been able to correlate the structural states of a single-domain signaling protein, NtrC, and its interconversion backbone dynamics on the microsecond time scale between active and inactive states with biochemical data (Buck and Rosen 2001). Using nuclear magnetic resonance (NMR) relaxation, Volkman et al. (2001) have characterized the motions of NtrC in three functional states: the inactive unphosphorylated state, the phosphorylated active state, and an unphosphorylated state of a mutant that is partially active. NtrC is part of a two-component system. It serves as a molecular switch, alternating between an inactive and active states. Volkman et al. (2001) have shown that first, in the millisecond-to-microsecond range, the unphosphorylated (inactive) forms were considerably more dynamic than the phosphorylated, active form. Second, these movements were in the same regions observed to undergo the largest conformational changes on phosphorylation. As expected, there was no difference in the fast nanosecond-to-picosecond side-chain motions between the two forms. Combined, these observations indicate that the conformational change relates to the difference in the active/inactive functional states. These results elegantly show that both conformational states (as well as many others) are populated by the dynamically fluctuating inactive unphosphorylated proteins. However, binding to the phosphate group leads to a population shift. Volkman et al. estimate the population of the active species within the population of the unphosphorylated molecules to be ~2% to 10%. Hence, the critical point here is that the phosphorylation did not actively induce a conformational change. Rather, it leads to a shift in pre-existing conformations, dynamically re-distributing the population. Moreover, mutants that are partially active without binding a phosphate group illustrate a similar microsecond shift as undergone by the wild-type unphosphorylated $NtrC^r$. Hence, mutations, like ligand binding, change the conditions, leading to population shifts. Figure 3 illustrates the hinge-bending motions in $NtrC^r$ and P-$NtrC^r$.

These results highlight two interesting points: First, despite the fact that some of the unphosphorylated $NtrC^r$ conformers pre-exist in an active-type conformation, no activity—in this case, basal transcription—is observed. Second, they indicate a mechanism through which a sharp switch can be accomplished. Following $NtrC^r$ phosphorylation, the molecule oligomerizes through a central domain-interaction. The activity as a function of the concentration is highly sigmoidal, with the population of the unphosphorylated $NtrC^r$ below the response threshold. This is in contrast to the behavior of CheY, which does not oligomerize following phosphorylation. CheY illustrates a low level of activity in the unphosphorylated state (Barak and Eisenbach 1992; Moy et al. 1994). In vivo, proteins are frequently found in

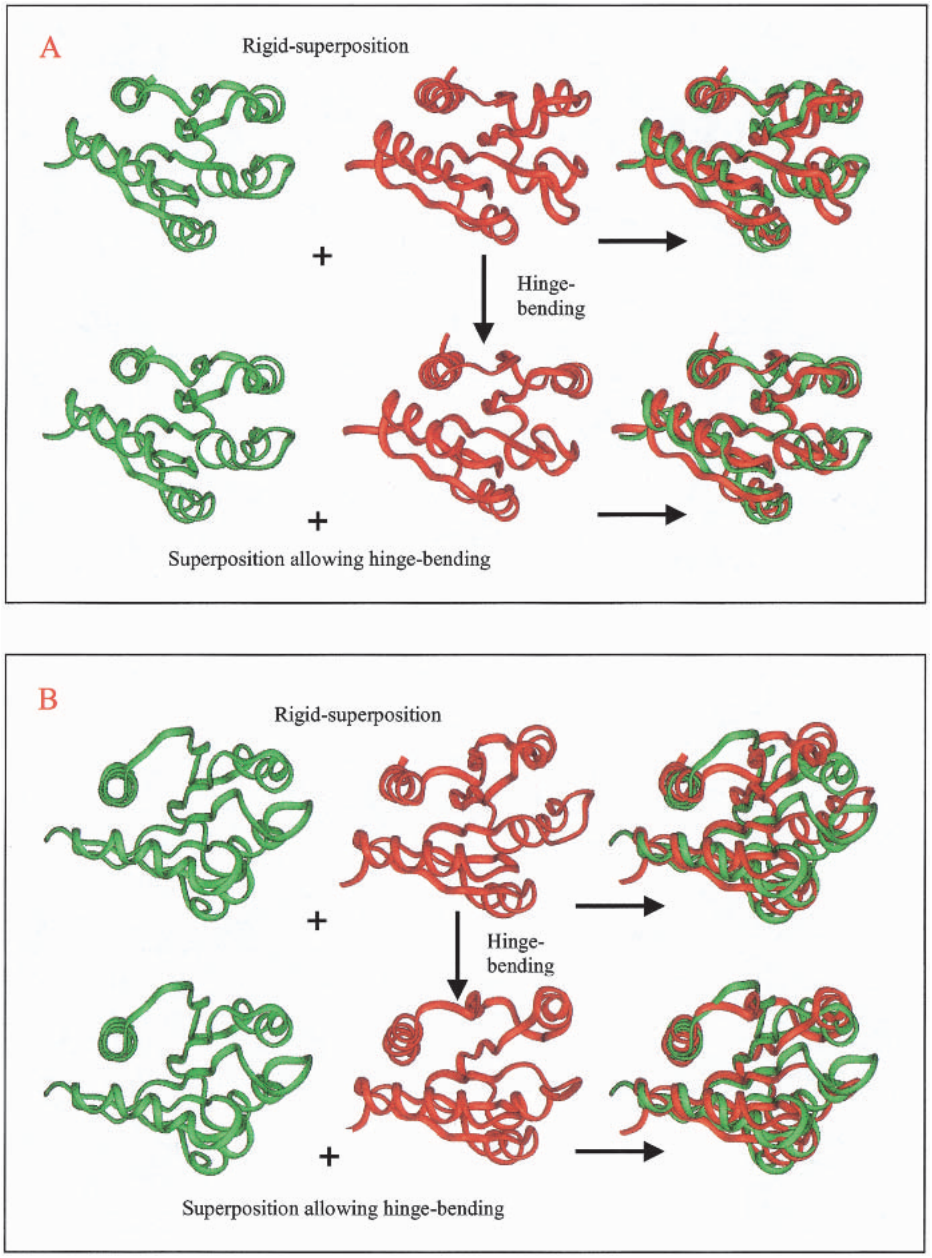

Fig. 3. Superposition of the inactive unphosphorylated state (NtrC′) and the phosphorylated active state (P-NtrC′). NtrC′ is the regulatory domain of NtrC, the nitrogen regulatory protein C, a signaling protein, acting as a molecular switch. The molecules have been superimposed using FlexProt, an automated, hinge-bending, flexible structural comparison algorithm (Shatsky et al. 2000; Shatsky 2001). (*A*) Comparison of the unphosphorylated state (green ribbon; Protein Data Bank [PDB] code, 1DC7; Bernstein et al. 1977) and the phosphorylated state (red ribbon; PDB code, 1DC8). (*B*). Comparison of the unphosphorylated state (green ribbon; PDB code, 1DC7) and another phosphorylated state (red ribbon; PDB code, 1NTR). The flexible structural comparison program can be accessed at http://bioinfo3d.math.tau.ac.il/FlexProt.

large multimolecular complexes, whether bound to sister molecules (in homo-oligomerization), to different molecules, or to both. To reach such a supramolecular assembly, proteins undergo a series of binding events, with a cascade of dynamic conformational redistributions. Each such event may play a role in regulation by modulating the conformational distribution. Such modulation offers evolutionary advantages and may be a function either of the barrier heights separating the conformers or of their difference in energies. Cascading redistributions will impact the likelihood of binding to multiple ligands of different shapes, sizes, and composition at a given site.

Above, we have addressed cases in which the native conformation is relatively highly populated. Here, the slow

hinge-bending motions occur between parts of folded proteins, even though the unbound unphosphorylated form nevertheless illustrates higher conformational variability at the binding site. That the binding site is still ordered is seen from the fact that Volkman et al. (2001) have been able to determine its atomic coordinates. However, some proteins, or protein parts, are natively disordered and are only observed ordered when bound to their ligands (Wright and Dyson 1999). In such proteins, when unbound, the disordered state is more stable than the native folded state, owing to a smaller extent of buried nonpolar surface area and (or) electrostatic repulsion in the native conformation (Uversky et al. 2000; Tsai et al. 2001). However, on binding, the buried nonpolar surface area at the intermolecular interface and the charge of the incoming ligand (likely the case in a phosphate group binding to $NtrC^r$ stabilize the native folded state. The issue of whether molecular disorder broadens or enhances ligand specificity has been controversial. In a disordered state, the population time of the native conformation is still higher than that of all other conformations; however, it is too low to be observed experimentally. Viewed in terms of populations, the relationship between disorder and broader or specific binding depends on the presence of ligands with binding that leads to favorable free energy that is significantly larger than the difference between the disordered and the native protein states.

The cascading population shifts in multiple-binding-site proteins, such as those in large multimolecular complexes, also explains the difference in binding (or dissociation) rates observed on increased ligand concentration in monovalent versus multivalent proteins (Rao et al. 1998; Buck and Rosen 2001). If the protein has a single binding site, binding is essentially the outcome of a collision-diffusion (and acceptance) process. Regardless of the concentration of the ligand, once the curve has reached a plateau, a further increase in ligand concentration will not change the binding rate (Rao et al. 1998), as the binding site is already saturated with colliding molecules. In contrast, in multiple-site molecules, as in those found in large complexes, the redistributions of the conformations affects the second (and third, etc.) sites, leading to enhanced binding rates at higher ligand concentrations.

Side-chain interactions across ensembles

Binding of ligands of different sizes, shapes, and composition implies different geometry of the receptor binding sites, movement of backbone and of side-chain atoms, and differences in the details of the residue interactions. Inspection of the interactions within the binding sites in structures of bound versus unbound molecules and comparisons of complexed structures with different ligands indicates that this is the case. Frequently, though not always, the same residues take part in the network of interactions. However, the way they interact differs, with different intermediary side-chain or backbone groups.

Here we present three types of examples to illustrate the variability in contacts in the same binding site of different receptor conformers. The first example concerns different ligands binding to a single receptor site (DeLano et al. 2000). Here all ligands are protein (peptide) molecules. The second example discusses the changes taking place in a receptor binding to one known ligand, an RNA molecule (Kranz and Hall 1999). Here, mutational analysis has been performed on the residues critical for binding in the binding site of this molecule. The analysis has illustrated how conserved and nonconserved regions of the mutant proteins can communicate (via cooperative interactions) to mediate RNA recognition. The third example illustrates the altered interactions across NMR conformer ensembles (Kumar and Nussinov 2001).

The first example concerns the Fc fragment (constant region) of immunoglobulin G complexed with its natural protein ligands—domain B1 of protein A, domain C2 of protein G, rheumatoid factor, and neonatal Fc-receptor—and with a combinatorially selected peptide, also interacting with the surface of IgG-Fc (DeLano et al. 2000). The peptide was observed to adopt a β-hairpin conformation unrelated to any known Fc binding motif. In spite of the absence of structural similarity between the peptide and the natural Fc protein ligands, the interactions formed between these and the Fc have many common features, including contacts with the polar residues Phe, Asn, Tyr, Asp, Glu, Lys, Gln, and others. All complexes show a shared set of contacts with side-chains and with backbone atoms, and an extensive nonpolar surface area. Nevertheless, the Fc binding site has slightly altered conformations in each of the complexes. The changes are largely observed in movements of two methionins to either form pockets (for protein G or for the peptide) or present a flatter surface (to the rheumatoid factor or to protein A). Ile, His, and Asn also adopt different rotameric states.

The second example relates to RNA recognition by the human U1A protein (Kranz and Hall 1999). The RNA-binding domains (RBD) are highly conserved, sharing a common α/β-sandwich tertiary fold, even though they bind diverse RNA ligands. The N terminus RNA-binding domain (RBD1) of the human U1A binds to stem/loop II of U1 snRNA, with high affinity. The β-sheet surface contains highly conserved solvent-exposed aromatic residues (Phe and Tyr in RBD1) and a Gln. NMR experiments have examined the effects of substitutions of these residues by similar residue types on the structure and dynamic properties. The results obtained by Kranz and Hall (1999) indicate that local interactions between these residues are mediated by conserved and nonconserved regions. Residue substitutions, like ligand binding, alter the distribution of the conforma-

tions, lead to population shifts, and enable glimpses of other conformations in solution.

The third example concerns the altered interactions observed in an ensemble of NMR conformers of 11 proteins, each containing at least 40 conformers (Kumar and Nussinov 2001). The analysis has focused on salt bridges occurring in the representatives of these ensembles (such as a crystal structure or an average energy-minimized structure, or if both are absent, of a most-representative structure) and its stabilizing/destabilizing contribution to protein stability in each conformer of the ensembles. This extensive electrostatic analysis of a total of 1249 ion pairs has shown that the vast majority of the ion pairs interconvert between being stabilizing and destabilizing to the structure at least once in the ensembles. These fluctuations reflect the variabilities in the location of the ion pairing residues and in the geometrical orientation of these residues, both with respect to each other and with respect to other charged groups in the rest of the protein. Additionally, salt bridges may exist in some conformers in the ensemble that do not exist in the representative structures, the outcome of side-chain movements.

Furthermore, comparison of proteins from different species, as well as mutational studies, have illustrated that apart from the critical importance of residues taking part in the catalytic mechanism, whereas there is a preference for conservation of some residues at the binding sites (Bogan and Thorn 1998; Hu et al. 2000), their presence is not an absolute requirement (Wells 1991; Clackson and Wells 1995). These studies have illustrated the relative resilience of binding sites to sequence alterations. They have further indicated altered hydrogen bonds between polar residue hot spots (Wedemayer et al. 1997; Kranz and Hall 1999; DeLano et al. 2000).

Two examples illustrate the variability in contacts in the same binding site by small molecule ligands. The first concerns pheromones binding to the mouse major urinary protein (MUP-I; Timm et al. 2001). These proteins function as carriers of volatile effectors of mouse physiology and behavior. Crystal structures of MUP-I complexed with two synthetic pheromones, 2-sec-butyl-4,5-dihydrothiazole (SBT) and 6-hydroxy-6-methyl-3-heptanone (HMH), have shown that the ligands define their orientations within the MUP-I β-barrel pocket. Timm et al. show that the ability of MUP-I to bind different lipophilic ligands derives from a limited extent of conformational flexibility and unoccupied space within the hydrophobic interior of the β-barrel. Interestingly, two water molecules present in the pocket hydrophobic environment enable at least two separate sites for polar groups in the pheromones to bind. Access to the barrel pocket is via flexible loops.

The second small molecule example relates to cAMP-dependent protein kinases (Engh et al. 1996) bound to H-series protein kinase inhibitors. H inhibitors bind in the adenosine binding sites of the protein kinases but do not bind to nucleotide binding proteins like actin, β-polymerase, or phosphofructokinase, probably owing to the lack of a complete binding pocket for the adenosine moiety. The triphosphate subsite appears to be highly conserved in protein kinases. On the other hand, a larger variability appears to exist in the residues lining the adenosine pocket. This subsite is the one used by the H inhibitors. The fact that some residues are involved in the H inhibitors binding to a larger extent than in purine binding helps to explain the larger variability in binding constants for some inhibitor classes as compared to binding to ATP.

Structural stability of binding sites

Englander and his colleagues (Milne et al. 1998, 1999) have proposed that proteins undergo local unfolding reactions throughout their structures. Although these may involve only a few residues, they may yield a large number of conformational states, in which each state is distinguished by the presence of locally unfolded region(s). This immediately indicates that the Gibbs energy of stabilization is not equally distributed in the protein structure. Some regions are more stable, and others are less stable. Freire and his colleague (Luque and Freire 2000) have consequently raised the question as to the biological implications of such uneven distributions of structural stability.

In particular, they have focused on protein binding sites. Shoichet et al. (1995) have shown that at least for T4 lysozyme, in the nine mutations at two substrate binding residues that they have studied, although the mutations increased the stability of the enzyme, they lowered its activity, implying that too much stability adversely affects enzyme function. Consistently, Streaker and Beckett (1999) have shown the importance of low-stability regions in allosteric proteins.

Freire and his colleagues have performed structure-based thermodynamic stability analysis using the COREX algorithm (Hilser and Freire 1996). Their results (Luque and Freire 2000) indicate that binding sites have a dual character. Binding sites appear to be characterized by the presence of regions of low stability and of high stability. The low stability is frequently, though not always, loops. Furthermore, in allosteric proteins, low-stability regions in regulatory sites appear to be critical for transmission of information to catalytic sites. Lower-stability regions exist in a range of conformational substates. However, we note that the motions studied by Freire are on a smaller scale, differing from the larger-scale hinge-bending motions. As such, although they may contribute to variable ligand/drug binding, the variablility is expected to be on a limited scale.

Implications for inhibitor design

Mutational studies have shown that there is a range of motions in the HIV protease and its retroviral homologs. A few

mutations, even though away from the binding site, may result in the pocket changing its volume and shape (Rose et al. 1998). Smaller drugs that initially fitted well might not be effective under such circumstances. Hence, although here the cause is the mutations and the end effect is drug resistance, owing to the change in the geometry of the binding site, it is via the hinge-bending swiveling domain (subdomain) flexibility that their effect may be transmitted. Because mutations do not create new conformations but lead to population shifts, such altered binding pockets exist in the native binding sites as well, indicating the need to consider bulkier (or smaller) drugs. By analyzing the range of motions observed in mutants and in variant structures, the extent of changes in size and volume might be inferred.

Furthermore, the recent demonstration of binding of multiple ligands of different sizes, shapes, and composition implies that higher affinity inhibitors may be designed. Currently, designed inhibitors are frequently modeled after the transition state of the substrate. If, however, the range of potential ligands is broader, dissimilar inhibitors may be derived. A good example is that of the *Lactobacillus casei* thymidylate synthase (LcTS; Tondi et al. 1999). The analogs found were dissimilar to the folate substrate but were shown to bind competitively with it. Here Tondi et al. have combined structure-based discovery with in-parallel synthetic techniques, which have enabled rapid elaboration of a series of compounds. Interestingly, the tighter binding inhibitors were also the most specific for LcTS compared with related enzymes.

On the other hand, if such multiple-drug single-site binding takes place primarily at sites governed by hinge joints, in which the connected parts can bend/rotate as rigid objects, constraining the hinge sites by protein design is a strategy to consider if higher affinity to a given ligand is sought. Here there are two considerations. The first relates to the residues at/near the hinge joints. Small residues, such as glycines, which allow too much flexibility, appear to be selected against (Sun and Sampson 1998), in addition to the presence of residues at the other end of the spectrum, which may lead to steric constraints. The second consideration is interdomain interactions. Inspection of the nonpolar buried surface area between domains, and between hinging subunits encompassing the active sites at their interface shows that it can be extensive. On the other hand, electrostatic interactions in the form of salt bridges seem to be discriminated against (N. Sinha, S. Kumar, and R. Nussinov, unpubl.).

Can we predict a likely extent of opening of the two domains with respect to each other? Clearly, the more extensive the nonpolar surface area between the two hinging domains in the closed conformation, the higher the contribution of the hydrophobic effect to the free energy. Because this contribution of the nonpolar buried surface area needs to be overcome if the two domains (or parts) swing out exposing the binding surface, this penalty needs to be overcome. Because the open conformation is present in the molecular ensemble before binding the ligand, overcoming the hydrophobic interactions that oppose the opening can most straighforwardly be performed via a compensating hydrophobic effect in the open conformation. Examination of closed and of their corresponding open conformations indicates that the open conformation can also bury an extensive extent of nonpolar surface area (N. Sinha, S. Kumar, and R. Nussinov, unpubl.). Interestingly, there appears to be an inverse correlation between the distance between the two opened domains compared with the distance in the closed conformation, and the extent of buried nonpolar surface area between the domains. The larger the extent of the nonpolar buried surface area in the closed conformation, the smaller the distance between the hinging parts in the open versus the closed conformation and, consequently, the larger the extent of the nonpolar buried surface area in the open conformation. Frequently, the magnitude of the buried surface is a function of the moving domain size. This indicates that if in the analyzed protein there is a substantial extent of nonpolar buried surface area between the domains in the closed conformation, we may expect a lesser extent of domain opening and, hence, pocket enlargement, affecting considerations relating to the corresponding drug design.

Flexible docking: Taking account of opulations

Computational docking is a frequently used tool in an effort to find lead inhibitors, new drugs, and prediction of modes of protein-protein associations. Frequently, the goal in docking is to start with a given enzyme/receptor and to search a large library, seeking potential high-affinity ligands. These may be of different sizes and composition. A major difficulty in docking schemes is the implementation of molecular flexibility. If the two molecules are taken as rigid, the magnitude of the problem is considerably reduced. However, molecules are flexible. In solution, docked configurations rejected owing to intermolecular penetrations may actually be accommodated with a favorable association, if we take into account molecular flexibility.

Yet, implementing flexibility in docking calculations is a major computational hurdle. Three major schemes have been adopted by researchers in this field (for review, see Nussinov and Wolfson 1999a,b; Nussinov et al. 2001). In the first, the ligand molecule is built into the active site, putting in a fragment at a time, allowing the connected fragments to rotate around their joint covalent bond. It can be applied only to ligands, taking the receptor as rigid. For even moderately sized ligands, such a procedure is computationally extremely expensive. The second approach is soft docking. This is a frequently adopted scheme. Here, rather than large-scale flexibility, one accounts only for surface plasticity. Consideration of surface atoms, or side-chain

movements, is implicitly taken care of via threshold intermolecular penetration penalty/reward schemes. The drawback of this approach is that it can account for only a limited surface plasticity. Allowing too large thresholds in such soft docking schemes will produce huge numbers of potential solutions and no clear way of ranking these. The third approach is to divide the ligand into parts, considering each part as a relatively rigid body and allowing swiveling hinge-bending motions between these. In such schemes, soft intermolecular penetration is frequently allowed. In this third hinge-bending category, three alternate computational schemes have been followed: In the first, one docks the first part (or domain), then allows all degrees of freedom for the second. This scheme implicitly entails a grid-based conformational search by the second domain and, hence, again, is too costly computationally. In the second scheme, one docks the parts separately, then seeks configurations in which the two parts are docked in consistent conformations, that is, in which they can be connected by a covalent bond with no intermolecular or interpart atomic overlaps. The drawback of such a scheme is the lack of utilization of information available a priori, namely, that the two parts are connected and that the location of the hinge is known. The third approach is the automated highly efficient robotics-based docking algorithm (Sandak et al. 1998, 1999). The performance of this method is attractive, with the parts docked simultaneously yet avoiding grid-based searches. Nevertheless, it is bounded by the problem of surface plasticity, thresholds, and the potential presence of too many hinges.

Docking of ligands of different sizes and shapes implicitly fits the concept of populations. The key is that there is no need to scan the entire conformational space in the docking computation. The conformational space is not randomly populated. Rather, focusing only on the more populated regions reduces the computational complexity. A practical scheme may involve three major steps. In the first, an ensemble of conformers is generated via (room temperature) simulations. These are subsequently clustered in a second step, and representatives are selected. The representatives are docked in the third step. By taking the representatives of clustered conformations, we implicitly focus on the more highly populated regions in space, rather than scanning it in its entirety. Within these, the more and less conserved regions may be scored appropriately. Knegtel et al. (1997) and Claussen et al. (2001) have already considered the ensemble concept in docking (Halperin et al. 2002). Figure 4 presents a schematic diagram of such a scheme.

Recently, Wrabl et al. (2001) have suggested assessing the goodness of fold recognition solutions through considerations of ensembles. Based on the notion of populations, they have used COREX (Hilser and Freire 1996) to compute preferences for amino acids to be in certain thermodynamic environments in the folded native structure. COREX is designed to reflect the local stability of the protein. It computes a thermodynamic ensemble-based description of the protein native state. The stability derives from the environment of the residue, whether a polypeptide segment or a building block of the structure. Remarkably, Wrabl et al. have shown that the amino acid types have different propensities to occur in high-, medium-, and low-stability environments within the protein. Based on these nonrandom distributions, they have derived three dimensional–to–one dimensional matrices for given proteins to evaluate the quality of predicted folded structures. Hence, here the idea has been to use the thermodynamic environment preferences of the residues in ensembles of native structures to rank fold recognition solutions. Wrabl et al. (2001) have further applied this novel population-based approach to a set of sequences to test the performance of such matrices compared with utilization of secondary structure information. Nevertheless, a potential concern is that in using such three dimensional–to–one dimensional matrices, the environmental information is lost. Hence, although utilization of such an ensemble approach to folding—and to binding, to assess docked configurations—is an attractive idea, one needs to design and implement a way of retaining the environmental information, which reflects the partitioning of the residues.

Conclusions: Specificity versus multiple diverse ligands binding at the same site

Throughout evolution, domain (subdomain) bending and rotation have been a simple solution for an organism/molecule to follow to achieve considerable molecular motions over large distances. Consistently, domain movements have long been shown to be involved in interconversion between closed versus open binding mechanism. Domain movements are directly involved in changes in the geometry of the binding site. Such movements may result in masking (or exposing) certain residues at the active site. The interrelationship between domain motions and the active site has been particularly well studied in cases such as in immunoglobulins, kinases, proteases, t-RNA synthetase, tryptophan synthase, and the GroEL molecular chaperone. In addition, examination of the combinatorially generated mutations in triose phosphate isomerase has illustrated trends in the requirements as to which residues can serve as adequate hinges. Further, mutations in one region of the protein may have effects at locations far away, via domain motions. Indeed, the most straightforward way to achieve large-scale movements is through such domain displacements.

Why then do the multiple ligands preferentially bind at a single site rather than at different locations on the receptor surface? A major factor appears to be the presence of potential hinges, in addition to energy hot spots. The concept of specificity has largely been inferred from natural ligand binding. However, given a large enough library of ligands,

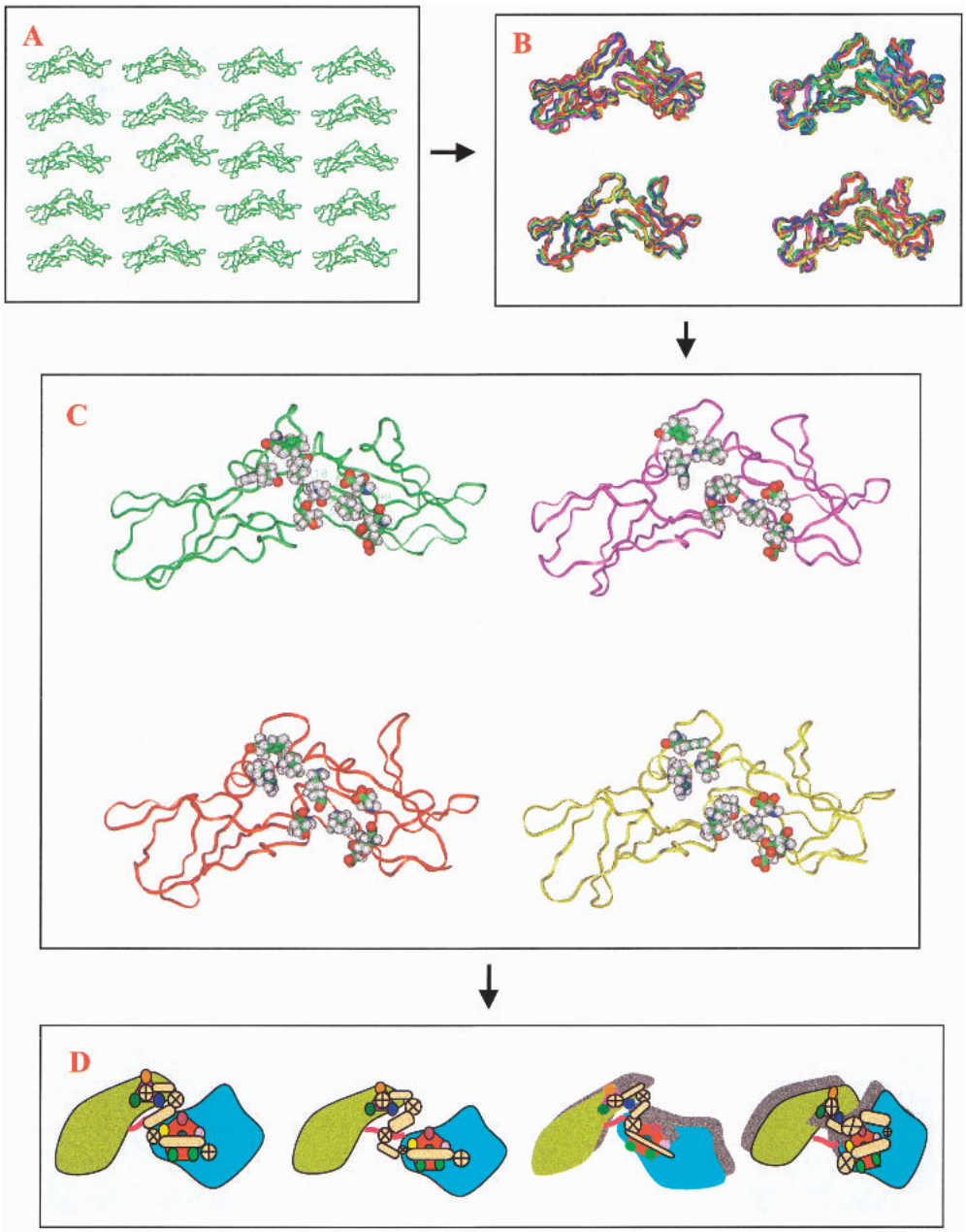

Fig. 4. A schematic diagram illustrating how taking account of populations may aid in flexible docking schemes. The protein illustrated is the tissue factor. Tissue factor is in complex with blood coagulation factor VIIa with the binding sites located in the hinge region (three residues on one domain, i.e., R131, L133, and F140, six residues on another domain, i.e., T17, K20, I22, E24, E56, and D58; Banner et al. 1996). (*A*) An ensemble of molecules is generated via molecular dynamics simulations. (*B*) The molecular ensemble is superimposed, generating some clusters. (*C*) Representatives are chosen for each large-enough cluster (the highlighted residues are the three residues on the first domain, and the six residues on the second domain noted above). (*D*) Ligands are docked into the representatives. See also Fig. 1.

it is reasonable to expect that additional high-affinity molecules will bind at the same site. The range of dissimilar ligands that will bind depends on the distribution and redistribution of the conformations, which relate to the flexibility at and around the hinge, on their free energy differences and on protein-ligand interactions.

The essential point is that molecules exist in dynamically changing populations, in equilibrium. Considering binding as a process, and that the binding pocket is defined by the incoming selected molecular partner, illustrates that specificity and multiple different-ligand single-site binding are essentially the same process. The principles are similar. The

process is guided by the receptor populations and by the ligands present in solution, via conformational selection.

Hence, identification of hinges and of the domains that move as rigid bodies with respect to each other on these hinges is essential. These may provide a rationale and guidelines for designing inhibitors that would fit into an altered larger binding site. For such a task, suitable and efficient computational approaches would be of considerable value.

Acknowledgments

We thank Drs. C.J. Tsai, S. Kumar, and, in particular, J.V. Maizel for discussions and encouragement. The research of R. Nussinov and H.J. Wolfson in Israel has been supported in part by the Magnet grant, by the Ministry of Science grant, and by the Center of Excellence in Geometric Computing and its Applications funded by the Israel Science Foundation (administered by the Israel Academy of Sciences). The research of H.J.W. is partially supported by the Hermann Minkowski-Minerva Center for Geometry at Tel Aviv University. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract No. NO1-CO-12400. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

References

Abad-Zapatero, C., Rydel, T.G., and Erickson, J. 1990. Revised 2.3 Å structure of porcine pepsin: Evidence of a flexible subdomain. *Proteins* **8:** 62–81.

Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins. J., Silva, A.M., and Erickson, J.W. 1993. Crystal structures of native and inhibited forms of human cathepsin D: Implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci.* **90:** 6796–6800.

Banner, D.W., D'Arcy, A., Chene, C., Winkler, F.K., Guha, A., Konigsberg, W.H., Nemerson, Y., and Kirchhofer, D. 1996. The crystal structures of the complex of blood coagulation factor VIIa with soluble tissue factor *Nature* **380:** 41–46.

Barak, R. and Eisenbach, M. 1992. Correlation between phosphorylation of the chemotaxis protein Che Y and its activity at the flaggelar motor. *Biochemistry* **31:** 1821–1826.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. 1977. The Protein Data Bank: A computer based archival file for macromolecular structures. *J. Mol. Biol.* **112:** 535–542.

Bogan, A.A. and Thorn, K.S. 1998. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280:** 1–9.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., Sate, D.J., Swaminathan, S., and Karplus, M. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* **4:** 187–217.

Buck, M. and Rosen, M.K. 2001. Flipping a switch. *Science* **291:** 2329–2330.

Carlson, H.A and McCammon, J.A. 2000. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **57:** 213–218.

Clackson, T. and Wells, J.A. 1995. A hot spot of binding energy in hormone-receptor interfaces. *Science* **267:** 383–386.

Clackson, T., Ultsch, M.H., Wells, J.A., and de Vos, A.M. 1998. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J. Mol. Biol.* **277:** 1111–1128.

Claussen, H., Buning, C., Rarey, M., and Lengauer, T. 2001. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308:** 377–395.

Cox, S., Radzio-Andzelm, E., and Taylor, S.S. 1994. Domain movements in protein kinases. *Curr. Opin. Struct. Biol.* **4:** 893–901.

Creighton, T.E. 1993. *Proteins: Structure and molecular properties*, 2nd ed. WH Freeman and Company, New York, NY.

DeLano, W.L., Ultsch, M.H., de Vos, A.M., and Wells, J.A. 2000. Convergent solutions to binding at a protein-protein interface. *Science* **287:** 1279–1283.

Demchenko, A.P. 2001. Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.* **14:** 42–61.

Engh, R.A., Girod, A., Kinzel, V., Huber, R., and Bossemeyer, D. 1996. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8 and H89: Structural implications for selectivity. *J. Biol. Chem.* **271:** 26157–26164.

Gerstein, M. and Krebs, W. 1998. A database of macromolecular motions. *Nucleic Acids Res.* **26:** 4280–4290.

Gerstein, M., Anderson, B.F., Norris, G.E. Baker, E.N., Lesk, A.M., and Chothia, C. 1993a. Domain closure in lactoferrin: Two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.* **234:** 357–372.

Gerstein, M., Schulz, G., and Chothia C. 1993b. Domain closure in adenylate kinase: Joints on either side of two helices close like neighboring fingers. *J. Mol. Biol.* **229:** 494–501.

Gerstein, M., Lesk, A.M., and Chothia, C. 1994. Structural mechanism for domain movements in proteins. *Biochemistry* **33:** 6739–6749.

Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. 2002. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins,* in press.

Hilser, V.J. and Freire, E. 1996. Structure based calculation of the equilibrium folding pathway of proteins: Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **262:** 756–772.

Horovitz, A. 1998. Structural aspects of GroEL function. *Curr. Opin. Struct. Biol.* **8:** 93–100.

Hu, Z., Ma, B., Wolfson H., and Nussinov, R. 2000. Conservation of polar residues as hot spots at protein–protein interfaces. *Proteins* **39:** 331–342.

Huang, M., Syed, R., Stura, E.A., Stone, M.J., Stefanko, R.S., Ruf, W., Edington, T.S., and Wilson, I.A. 1998. The mechanism of an inhibitory antibody on TF initiated blood coagulation revealed by the crystal structures of human tissue factor FAB 5G9 and TF-FAB 5G9 complex. *J. Mol. Biol.* **275:** 873–894.

Huber, R. 1987. Flexibility and rigidity, requirements of the function of proteins and protein pigment complexes. *Biochem. Soc. Trans.* **15:** 1009–1020.

Janin, J. and Wodak, S. 1983. Structural domains in proteins and their role in the dynamics of protein function. *Prog. Biophys. Mol. Biol.* **42:** 21–78.

Joseph, D., Petsko, G.A., and Karplus, M. 1990. Anatomy of conformational change: Hinged lid motion of the trisephosphate isomerase loop. *Science* **249:** 1425–1428.

Keskin, O., Jernigan, R.L., and Bahar, I. 2000. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.* **78:** 2093–2106.

Knegtel M.A., Kuntz, I.D., and Oshiro, C.M. 1997. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **266:** 424–440.

Koshland, D.E. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci.* **44:** 98–123.

Kranz, J.K. and Hall, K.B. 1999. RNA recognition by human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface. *J. Mol. Biol.* **285:** 215–231.

Kumar, S. and Nussinov, R. 2001. Fluctuations in ion pairs and their stabilities in proteins. *Proteins* **41:** 485–497.

Kumar, S., Ma, B., Tsai, C.J., Sinha, N., and Nussinov, R. 2000. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci.* **9:** 10–19.

Lazaridis, T. and Karplus, M. 1999. Effective energy function for protein in solution. *Proteins* **35:** 133–152.

Lee, A.Y., Gulnik, S.V., and Erickson, J. 1998. Conformational switching in an aspartic proteinase. *Nat. Struct. Biol.* **5:** 866–871.

Luque, I. and Freire, E. 2000. Structural stability of binding sites: Consequences for binding affinity and allosteric effects. *Proteins* **4:** 63–71.

Ma, B., Kumar, S., Tsai, C.J., and Nussinov, R. 1999. Folding funnels and binding mechanisms. *Protein Eng.* **12:** 713–720.

Ma, B., Wolfson, H., and Nussinov, R. 2001. Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr. Opin. Struct. Biol.* **11:** 364–369.

Mattevi, A., Rizzi, M., and Bolognesi, M. 1996. New structures of allosteric proteins revealing remarkable conformational changes. *Curr. Opin. Struct. Biol.* **6:** 824–829.

Milne, J. S., Mayne, L., Roder, H., Wand, A.J., and Englander, S.W. 1998. Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Sci.* **7:** 739–745.

Milne, J.S., Xu, Y., Mayne, L.C., and Englander, S.W. 1999. Experimental study of the protein folding landscape: Unfolding reactions in cytochrome c. *J. Mol. Biol.* **290:** 811–822.

Moy, F.J., Lowry, D.F., Matsumura, P., Dahlquist, F.W., Krywko, J.E., and

Domaille, P.J. 1994. Assignments, secondary structure, global fold, and dynamics of chemotaxis Y protein using three- and four-dimensional heteronuclear (13C,15N) NMR spectroscopy.. *Biochemistry* **33:** 10731–10742.

Muller, Y., Kelley, R.F., and De Vos, A.M. 1998. Hinge bending within the cytokine receptor superfamily revealed by the 2.4 Å crystal structure of the extracellular domain of the rabbit tissue factor. *Protein Sci.* **7:** 1106–1115.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247:** 536–540.

Nussinov, R. and Wolfson, H. 1999a. Efficient computational algorithms for docking, and for generating and matching a library of functional epitopes, I: Rigid and flexible hinge-bending docking algorithms. *Comb. Chem. High Throughput Screen.* **2:** 249–259.

———. 1999b. Efficient computational algorithms for docking, and for generating and matching a library of functional epitopes, II: Computer vision-based techniques for the generation and utilization of functional epitopes: Rigid and flexible hinge-bending docking algorithms. *Comb. Chem. High Throughput Screen.* **2:** 261–269.

Nussinov, R., Ma, B., and Wolfson, H. 2001. Computational methods for docking and applications to drug design: Functional epitopes and combinatorial libraries. In *Current topics in computational biology* (J. Jiang et al., eds). MIT Press, Boston, MA. (in press).

Oh, B.H., Pandit, J., Kang, C.H., Nikaido, K., Gokcen, S., Ames, G.F.L., and Kim, S.H. 1993. Three-dimensioanl structures of the periplasmic lysine/arginine/ornithine-binding proteins with and without a ligand. *J. Biol. Chem.* **268:** 11348–11355.

Rao, J., Joydeep, L., Isaacs, L., Weis, R.M., and Whitesides, G.M. 1998. A trivalent system from vancomycin-D-ala-D-ala with higher affinity than avidin-biotin. *Science* **280:** 708–711.

Ringe, D. 1995. What makes a binding site a binding site? *Curr. Opin. Struct. Biol.* **5:** 825–829.

Rose, R.B., Craik, C.S., and Stroud, R.M. 1998. Domain flexibility in retroviral proteases: Structural implications for drug resistant mutations. *Biochemistry* **37:** 2607–2621.

Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T., and Blundell, T.L. 1992. Domain flexibility in aspartic proteinases. *Proteins* **12:** 158–170.

Sandak, B., Wolfson, H.J., and Nussinov, R. 1998. Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *Proteins* **32:** 159–174.

Sandak, B., Nussinov, R., and Wolfson, H.J. 1999. A method for biomolecular structural recognition and docking allowing conformational flexibility. *J. Comput. Biol.* **5:** 631–654.

Schulz, G.E., Muller, C.W., and Diederichs, K. 1990. Induced-fit movements in adenylate kinases. *J. Mol. Biol.* **213:** 627–630.

Sharff, A.J., Rodseth, L.E., Spurlino, J.C., and Quiocho, F.A. 1992. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31:** 10657–10663.

Shatsky M. 2001. "Alignment of flexible protein structures." M.Sc. thesis, Tel Aviv University.

Shatsky, M., Fligelman, Z., Nussinov, R., and Wolfson, H. 2000. Alignment of flexible protein structures. Proceedings of the 8th Conference on Intelligent Systems in Molecular Biology (ISMB) (In R. Altman et al., eds.), pp. 329–343. AAAI Press, Menlo Park, CA.

Shoichet, B.K., Baase, W.A., Kuroki, R., and Mattews, B.W. 1995. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci.* **92:** 452–456.

Silva, A.M., Lee, A.Y., Gulnik, S.V., Maier, P., Collins, J., Bhat, T.N., Collins, P.J., Cachau, R.E., Luker, K.E., Gluzman, I.Y., Francis, S.E., Oksman, A., Goldberg, D.E., and Erickson, J.W. 1996. Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from *Plasmodium falciparum.* *Proc. Natl. Acad. Sci.* **93:** 10034–10039.

Sinha, N. and Nussinov, R. 2001. Point mutations and sequence variability in proteins: Re-distributions of pre-existing populations. *Proc. Natl. Acad. Sci.* **98:** 3139–3144.

Streaker, E.M. and Beckett, D. 1999. Ligand-linked structural changes in the *Escherichia coli* biotin repressor: The significance of surface loops for binding and allostery. *J. Mol. Biol.* **292:** 619–632.

Sun, J. and Sampson, N.S. 1998. Determination of the amino acid requirements for a protein hinge in triose phosphate isomerase. *Protein Sci.* **7:** 1495–1505.

Sundberg, E.J. and Mariuzza, R.A. 2000. Luxury accommodations: The expanding role of structural plasticity in protein-protein interaction. *Structure* **8:** R137–R142.

Timm, D.E., Baker, L.J., Mueller, H., Zidek, L., and Novotny, M.V. 2001. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Sci.* **10:** 997–1004.

Tondi, D., Slomczynska, U., Costi, M.P., Watterton, D.M., Ghelli, S., and Shoichet, B.K. 1999. Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chem. Biol.* **6:** 319–331.

Tsai, C-J., Ma, B., and Nussinov, R. 1999. Folding and binding cascades: Shifts in energy landscapes. *Proc. Natl. Acad. Sci.* **96:** 9970–9972.

Tsai, C.J., Ma, B., Sham, Y., Kumar, S., and Nussinov, R. 2001. Structured disorder and conformational selection. *Proteins* **44:** 418–427.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41:** 415–427.

Van Regenmortel, M.H.V. 1999. Molecular recognition in the post-reductionist era *J. Mol. Recognit.* **12:** 1–2.

Vazquez-Laslop, N., Zheleznoca, E.E., Markham, P.N., Brennan, R.G., and Neyfakh, A.A. 2000. Recognition of multiple drugs by a single protein: A trivial solution of a old paradox. *Biochem. Soc. Trans.* **28:** 517–520.

Volkman, B.F., Lipson, D., Wemmer, D.E., and Kern, D. 2001. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291:** 2429–2433.

Wedemayer, G., Patten, P., Wang, L., Schultz, P., and Stevens, R. 1997. Structural insights into the evolution of an antibody combining site. *Science* **276:** 1665–1669.

Wells, J.A. 1991. Systematic mutational analyses of protein-protein interfaces. *Methods Enzymol.* **202:** 390–411.

Wrabl, J.O., Larson, S.A., and Hilser, V.J. 2001. Thermodynamic propensities of amino acids in the native state ensemble: Implications for fold recognition. *Protein Sci.* **10:** 1032–1045.

Wright, P.E. and Dyson, H.J. 1999. Intrinsically unstructured proteins: Reassessing the protein structure-function paradigm. *J. Mol. Biol.* **293:** 321–331.

Zwahlen, C., Li, S.C., Kay, L.E., Pawson, T., and Forman-Kay, J.D. 2000. Multiple modes of peptide recognition by the PTB domain of the cell fate determinant Numb. *EMBO J.* **19:** 1505–1515.